

## Fe de erratas / Erratum

Medea-García, Lucía, “*Ḥattā* throughout the centuries: origin and history of this particle in Classical Arabic”, *Al-Qaṭara* 39, 2 (2018), pp. 503-545.  
<https://doi.org/10.3989/alqantara.2018.015>

- Se han corregido los apellidos de la autora (“Medea-García” en lugar de “García-Medea”) en la primera página y en el encabezamiento de las páginas pares.

- The last name of the author has been corrected (“Medea-García” instead of “García-Medea”) in the first page and header of even pages.

6 de junio de 2019

## ***Ḥattā* throughout the centuries: origin and history of this particle in Classical Arabic**

*Ḥattā* a lo largo de los siglos:  
origen e historia de esta partícula en árabe clásico

Lucía Medea-García  
Universidade de Lisboa  
<http://orcid.org/0000-0002-2946-3370>

The main objectives of this article are, on the one hand, to propose a methodology for the study of language change in Classical Arabic, and, on the other hand, to explore the specificities of grammaticalization and language change in this Semitic language. This work analyses the processes of evolution and grammaticalization undergone by the particle *ḥattā* ('until', 'even') in Classical Arabic between the 7th and the 20th centuries. 731 examples of *ḥattā* in *fuṣṣḥā* have been examined, using one of the most complete historical corpora of the Arabic language, the HADC, still unpublished. The study considered 30 different variables that have been organized in a relational database. The results are presented in a conceptual map showing the semantic bonds of the five values of *ḥattā*, as well as the evolutionary pathways of this particle in Classical Arabic.

The main conclusions are that *fuṣṣḥā* does indeed undergo language change; its evolution is, however, constrained by the Quranic norm and its attachment to the written form. Results also reveal some of the specificities of grammaticalization in Classical Arabic, such as the marked *layering*, *persistence* and *divergence*, or the relevance of including variables such as the origin of the author or the types of text when dealing with language change.

Los objetivos principales de este trabajo son, por una parte, plantear una propuesta metodológica para el estudio del cambio lingüístico en árabe clásico y, por otra, explorar las particularidades de los procesos de gramaticalización y el cambio lingüístico en esta lengua semítica. Para ello, hemos analizado los procesos de cambio y gramaticalización experimentados por la partícula *ḥattā* ('hasta', 'incluso') en árabe clásico desde el siglo VII hasta el XX. Se han analizado 731 ejemplos de *ḥattā* extraídos de uno de los más completos corpus históricos del árabe, el HADC, todavía inédito. El estudio ha tenido en cuenta 30 variables diferentes que se han recogido en fichas recopiladas en una base de datos relacional. Los resultados se han organizado en un esquema conceptual que presenta la vinculación semántica de cada uno de los cinco valores de *ḥattā* así como las trayectorias evolutivas de esta partícula en árabe clásico.

Las conclusiones más relevantes son que, por descontado, sí hay cambio lingüístico en árabe clásico, si bien sus trayectorias evolutivas se ven constreñidas tanto por la presión de la norma coránica como por su naturaleza de lengua esencialmente escrita. Los resultados también revelan que existe una marcada *estratificación*, *persistencia* y *divergencia* en los procesos y apuntan, a su vez, la relevancia de incluir

*Key words:* grammaticalization; *hattā*; prepositions; corpus linguistics; cognitive linguistics; Classical Arabic.

variables como el origen del autor o los tipos de texto cuando se analiza el cambio lingüístico.

*Palabras clave:* gramaticalización, *hattā*, preposiciones, lingüística de corpus, lingüística cognitiva, árabe clásico.

## 1. Introduction

The Arabic language has an extensive and exceptionally rich literary and grammatical tradition; it has not, however, been the subject of much diachronic research. This study has two main objectives. On the one hand, it attempts to provide a rigorous methodology for the study of language change in Classical Arabic<sup>1</sup>; on the other hand, it aims to explore the specificities of grammaticalization processes and language change in *fuṣḥā* through the case of the particle *hattā*.

The decision of doing research in the area of language change in Arabic comes motivated by two main reasons. The first is the astonishing lack of studies dedicated to this domain of study, due in great measure to the different factors that hinder it from developing. Classical Arabic is constrained by certain ideological, religious and political beliefs which hold that it is historically fixed and essentially unaffected by language change<sup>2</sup>. This reasoning is due to the sacred nature often attributed to it, which argues that it mainly originated as the language of the Quran, which is simultaneously its greatest expression. *Fuṣḥā* is also a symbol of pan-Arabic unity, and some consider that allowing it to change would threaten this unity. These factors, along with the lack of linguistic tools such as corpora, historical grammars and dictionaries, and the absence of a strong research tradition in which to ground it,

<sup>1</sup> The terminology used in this article will be the following: “Classical Arabic” and “*fuṣḥā*” will be used as synonyms, meaning the written Arabic used in the literary and scientific tradition. The more specific denomination “Modern Standard Arabic” will designate the current standard Arabic, used in written and formal oral contexts from the 19th century (in the academic domain, in the media, etc.). Finally, we will use the terms “Arabic dialects” and “vernacular languages” as synonyms.

<sup>2</sup> Versteegh, “The Arabic Language”, and Haeri, “Sacred language, ordinary people”.

make the diachronic research of this Semitic language extremely difficult to accomplish.

The second reason that has incited us to do research in the branch of historical linguistics in Arabic has been, precisely, the characteristics of Arabic itself. This language has developed a continuous and broad written tradition going back to the 4th century that can be traced up until today. It covers a wide geographical area and boasts one of the oldest and most prolific grammatical traditions. Furthermore, “there is nowhere in language history where members of the same family can be compared over such a long period of time as with the Semitic family”<sup>3</sup> and it is also “the Semitic language with the largest attested lexical inventory”<sup>4</sup>. The Arabophone world covers a wide area, from Mauritania to Oman, so this language also represents a privileged *locus* for the analysis of change triggered by contact language and areal grammaticalization, among others. And all this, without mentioning the fact that it comprises dozens of dialects that can provide information about previous stages of the language<sup>5</sup>. Indeed, the characteristics of this Semitic language turn it into a stimulating object of study for diachronic approaches.

In order to delve into the history of Arabic, we have explored and followed the trajectory of the element *ḥattā* (‘until’, ‘even’) from the 7th until the 20th century. The choice of this particle responds to different factors. It is a very ancient and, at the same time, very alive and currently used word: we find it in Arabic texts from the 6th century<sup>6</sup>, but it is also used nowadays in *fushḥā* and in all Arabic dialects. From its beginnings it was a polyfunctional and polysemic element, and the hypotheses about its etymological origin were, at least until 2017<sup>7</sup>, quite obscure. There is another noteworthy aspect of *ḥattā*: it has been adopted by at least twelve languages from four different language fam-

<sup>3</sup> Owens, “A linguistic history”, p. 30.

<sup>4</sup> Rendsburg, Rubin and Huehnergard, “A proper view of Arabic”, p. 538.

<sup>5</sup> Owens, “A linguistic history”, pp. 8-13.

<sup>6</sup> *Ḥattā* appears in five of the seven *Mu‘allaqāt*; concretely, in the poems of ʿArafa, Labīd, ‘Amr Ibn Kulthūm, ‘Antara Ibn Šaddād and al-Ḥārith Ibn Ḥilliza.

<sup>7</sup> Indeed, last year (2017) appeared the work “Dictionnaire du faisceau dialectal” by Corriente, Pereira and Vicente, as well as the PhD dissertation by the author of this article; and this year, 2018, the article “The origins of *ḥattā*”, by Al-Jallad, was published: all these works offer more convincing hypotheses (than the traditional ones) about the origins of *ḥattā*.

ilies<sup>8</sup>. Such a productive and long-standing word seemed the perfect element for our diachronic study.

This article can be divided in six sections: after this introduction, we will gather, in section 2, the main hypothesis about the etymological origin of *ḥattā*, as well as establish our own proposal. Section 3 will be devoted to the methodology, and in section 4 we will analyse the evolution of the semantic values of *ḥattā* through the centuries. A cognitive approach to complement this analysis, together with a typological contrast, will be offered in section 5, before ending with the conclusions.

## 2. Etymological origin of *ḥattā*

The etymological origin of *ḥattā* is quite obscure or, as Langer<sup>9</sup> claims, unknown. There are three main hypotheses regarding this issue. The first one includes the several postures defended within the Arabic Grammatical Tradition (henceforth, AGT) as well as by some contemporary grammarians<sup>10</sup>. We could say that this hypothesis claims, ultimately, that *ḥattā* comes from the verbal root *ḥtt*. There are several opinions, however, regarding the meaning of the verb derived from *ḥtt*. Some defended that it means ‘he scraped it, or rubbed it, off [...] namely, a thing, as, for instance, blood, or semen, or something dry, from a garment, or the like’<sup>11</sup>, while others claim it means ‘quickness’ or ‘ending something’<sup>12</sup>. The preposition *ḥattā* would be semantically connected to this verb derived from *ḥtt* through the idea of an ending point, for example, scraping something “until the end”<sup>13</sup>.

From a different perspective, Corriente, Pereira and Vicente<sup>14</sup> link the origins of *ḥattā* with the pan-Semitic root ‘*d*. From this root we would also have the Hebrew and Aramaic form ‘*ad*, the element ‘*d*

<sup>8</sup> As far as we know, *ḥattā* has been adopted at least by languages such as Spanish, Portuguese, Galician, Farsi, Swahili, Turkish, Turkmen, Urdu, Kurdish, Kazakh, Azerbaijani and Siwi.

<sup>9</sup> Langer, “Conjunctions”, p. 467.

<sup>10</sup> For example, Essees, “Grammaticalization of Arabic”, p. 320.

<sup>11</sup> Lane, “An Arabic-English Lexicon”, p. 508b.

<sup>12</sup> Sadan, Arik. “The subjunctive”, p. 199.

<sup>13</sup> Sadan, Arik. “The subjunctive”, p. 200.

<sup>14</sup> Corriente, Pereira and Vicente, “Dictionnaire du faisceau dialectal”, p. 312.

from the Ugaritic and the Epigraphic Sudarabic, *‘adi* from Ge’ez and, finally, the Acadian form *adi*. The authors suggest that *ḥattā* could originate from the union between this pan-Semitic root with the feminine deictic element *\*taī* > *\*attaī*. They propose, besides, a combination with (a part of) the word *ḥaītu* ‘where’, in order to explain the desonorization of the present form (that exhibits an initial *ḥ*). The process would be as follows: *\*ḥaī(tu) + ‘ad+taī* > *\*ḥa + ‘attaī* > *ḥattā*.

Finally, there is a third proposal by Al-Jallad<sup>15</sup>, who defends that *ḥattā* comes from the union between the noun *ḥadd* (‘limit’) and the subordinating conjunction *\*taī*: *ḥadd + taī* > *\*ḥattaī* > *ḥattā*.

*Ḥattā* appears to be an element with uncertain origins despite the attention received over the time. From our point of view, the etymological origin of *ḥattā* is linked to the noun *ḥadd* (‘limit’) and, at the same time, both elements are related to the pan-Semitic root *‘d*. Firstly, we will expose the reasons why *ḥattā* could come from the element *ḥadd*; after that, we will expand on the relationship of these two words with the root *‘d*.

### 2.1. Links with the noun *ḥadd* (‘limit’)

(i) Firstly, from a cognitive point of view, it seems logic to connect a spatial noun such as ‘limit’ with a conjunction or preposition that codifies the value ‘spatiotemporal limit’ such as *ḥattā*. Indeed, one of the main sources in the development of adpositions are spatial nouns. (ii) Second, the links of *ḥattā* (mainly a temporal element nowadays) with the noun *ḥadd* (completely spatial) responds to the very known grammaticalization chain SPACE > TIME, that is, the category TIME can be expressed in spatial terms. (iii) In third place, both elements are, phonetically, very similar. *Ḥadd* ends with a double *d* that, in the particle *ḥattā*, would appear voiceless: *ḥadd* > *ḥatt*. A possible path to the form would have been the addition of a final vocal support, generally carried out by the neutral vowel *schwa*. In *ḥattā* this *schwa* would be codified by the letter *‘ālif maqṣūra*, maybe by analogy with other two-syllabled Arabic prepositions, such as *‘alā* and *‘ilā*. Another possible

<sup>15</sup> Al-Jallad, “The Etymology of *Ḥattā*”.

explanation is given by Al-Jallad<sup>16</sup>: the subordinating conjunction—or, in Corriente, Pereira and Vicente's words<sup>17</sup>, “the feminine deictic element”—\**taī*/\**tī*, would have been added to the noun *ḥadd*: *ḥadd* + *taī* > *ḥattāi* > *ḥattā*. (iv) In fourth place, some current Levantine dialects use the form *la-ḥadd* precisely as a preposition and conjunction (‘until’) that codifies ‘spatiotemporal limit’. What is more, even in *fushḥā* we find examples of the expression *’ilā ḥadd* used exactly as *ḥattā*<sup>18</sup>. (v) Finally, another of the reasons to defend this etymological origin of *ḥattā* is the fact that this same evolutionary path has been found in other (and very different) languages. Concretely, Heine and Kuteva<sup>19</sup> point out an example in Swahili: “*m-paka* ‘border’, ‘boundary’, noun > (*m*)*paka* ‘until’, temporal preposition, conjunction”, a path that can be also found in the Moré language, both of them belonging to the Niger-Congo family. Likewise, we can add examples of similar paths experienced by other languages: Catalan *fins*, or Italian *fino* (‘until’) are both derived from Latin *fine* ‘limit’<sup>20</sup>. This typological contrast constitutes another argument in favour of the evolutionary pattern we defend.

## 2.2 Links with the root ‘d

The 13th century lexicographer Ibn Manẓūr explained, in his *Lisān al-‘arab*, that some of the pre-Islamic tribes pronounced *ḥattā* as *‘attā*<sup>21</sup>, information also provided by the 13th century grammarian Irbillī<sup>22</sup>. Indeed, Ibn Manẓūr described that many people, including Ibn Mas‘ūd, one of the most important companion of the prophet Muhammad, read the Quranic expression *‘attā ḥīn* (instead of *ḥattā ḥīn*). Wright not only gathers this information, but he also points out that

<sup>16</sup> Al-Jallad, “The Etymology of *Ḥattā*”.

<sup>17</sup> Corriente, Pereira and Vicente, “Dictionnaire du faisceau dialectal”, p. 312. The translation is mine.

<sup>18</sup> Dozy, in his “Supplément aux dictionnaires arabes”, p. 255, offers an example in which *’ilā ḥadd al-mawt* is translated as follows: “[ils les frappèrent], jusqu’à ce qu’ils fussent près de mourir”.

<sup>19</sup> Heine and Kuteva, “World Lexicon”, p. 75.

<sup>20</sup> Visconti, in “On the origins of scalar”, p. 244, analyses the case for Italian.

<sup>21</sup> Ibn Manẓūr, “*Lisān al-‘arab*”, p. 43. Precisely, he claims that the form *‘attā* was used by the Hudhail and Zaqif tribes.

<sup>22</sup> Irbillī, “*Kitāb Ġawāḥir*”, p. 404.

this dialectal variant, *‘attā*, “may perhaps be etymologically connected with the corresponding Hebrew word *‘d*”<sup>23</sup>. However, and as Corriente, Pereira and Vicente<sup>24</sup> defend, this variant would be etymologically connected not only with the Hebrew ‘until’ (*‘d*), but with the pan-Semitic root *‘d*, also found in Ugaritic, Acadian, Aramaic, Epigraphic Sudarabic and Ge’ez. Likewise, we believe that this root *‘d* is also linked to the Safaitic element *ḥt*, translated by Al-Jallad as ‘until’<sup>25</sup>. We could conclude that there are three variants that come from the pan-Semitic root *‘d*: some of them start with the letter *‘*, such as the Hebrew, Aramaic, Ugaritic, Epigraphic Sudarabic and Ge’ez forms. Other languages would have adopted the form with an initial *h*, such as the Arabic (although with a variant with an initial *‘*, as pointed out before), and Safaitic. Finally, we would find languages in where this first phoneme would have disappeared and the word would start directly with a vowel, as in Acadian *adi*; and we could also mention Modern Hebrew, where this initial *‘* is not pronounced and the resulting word for ‘until’ is *ad*. These data indicates that the alternation or evolution from *‘* to *h*, or even to *Ø*, at least in initial position, is possible in Semitic languages.

To briefly sum up, we defend that the closest origin of *ḥattā* could be found in the Arabic noun *ḥadd* ‘limit’ and that, at the same time, the etymological origin of both elements (*ḥattā* and *ḥadd*) is linked to the pan-Semitic root *‘d*. This proposal would combine the Corriente, Pereira and Vicente’s hypothesis with Al-Jallad’s<sup>26</sup>. In any case, it is worth noting how complex is to be conclusive when dealing with etymologies and such ancient elements.

### 3. Methodology

Tracing the history of a language and its processes of change is, without doubt, a very laborious and complex task. In the case of Arabic, this mission faces even more obstacles than those found when dealing with European languages. In Arabic we do not have a strong research

<sup>23</sup> Wright, “A grammar of the Arabic”, p. 147.

<sup>24</sup> Corriente, Pereira and Vicente, “Dictionnaire du faisceau dialectal”.

<sup>25</sup> Al-Jallad, “An outline of the grammar”, p. 164.

<sup>26</sup> Corriente, Pereira and Vicente, “Dictionnaire du faisceau dialectal” and Al-Jallad, “The Etymology of *Ḥattā*”.



tradition in this field, and there are fewer tools necessary for a diachronic approach, such as etymological dictionaries, historical corpora or automatic morphosyntactic analysers, among others. Besides, language change can also be a controversial field because of religious and sociopolitical issues. For all these reasons, it is crucial to develop a solid and rigorous methodology that allows us to overcome, to the extent possible, some of these difficulties, and to provide reliable results.

Our methodological proposal to study language change in Classical Arabic can be divided in three main sections. (i) In the first place, we have established the “macrovariables”, that is, we have defined several identifying variables in order to carry out the selection of texts to study. (ii) Secondly, we undertook the task of finding an appropriate historical corpus that allowed us to work with all our previously selected variables. (iii) Finally, we determined the “microvariables”, that is, linguistic variables that help us gather morphosyntactic, semantic and even pragmatic information to examine in detail each case of study. In order to systematize the grouping and search of all these data we have designed an analysis datasheet and a relational database.

### 3.1. *Macrovariables*

We have established four macrovariables that will allow us to generate a controlled selection of texts, from which we will obtain the examples of *ḥattā* to analyse. These four variables are the following: (i) *type of text*; (ii) *time* (the historical period of the studied texts); (iii) *place* (the geographical origin of the author) and, finally, (iv) *type of language* (a relevant characteristic given the diglossic situation of Arabic). Each of these variables affects the selection of the other variables and, at the same time, is conditioned by them.

(i) *Type of text*. The forms and contents of specific type of texts generate different expressive needs, and these genre conventions are often linked to processes of language change.<sup>27</sup> Therefore, we need to control the type of text under study if we are to develop a comprehensive analysis of grammaticalization processes. For this research, historical-geo-

<sup>27</sup> Kabatek, “Tradiciones discursivas jurídicas”; Kabatek, “Tradiciones discursivas y cambio lingüístico”.

graphical texts have been selected, a category in which we include four very interrelated genres: historical texts, geographical texts, genealogies and travel literature. All of them have in common the presence and description of spatiotemporal aspects.

Five arguments justify the election of this type of text. a) We have examples of these genres continuously since the 9th century; in fact, it is the only type of text<sup>28</sup> uninterruptedly found over the centuries in the historical corpus used for this study<sup>29</sup>. b) Secondly, writings characterized as historical are very relevant to the study of a spatiotemporal particle such as *ḥattā*. They allow us analysing the behaviour of this element in more supposedly neutral texts, such as the historical or geographical documents, but also in more literary environments, such as the travel literature. c) Another reason to choose this group of texts is the fact that they are easily classifiable and usually written in prose, and they enjoy a long and prestigious tradition in the Arab world. d) In fourth place, we also selected this kind of compositions because, as Versteegh points out, they are often written in a “lower style” that we find “in private letters and in non-literary writing, such as geographical works, historiography...”<sup>30</sup>. In this kind of works “we find a relaxation of the strict standards, the introduction of colloquialisms and a businesslike style”<sup>31</sup> and, therefore, it exhibits a more interesting type of language for studies about evolution and language change. e) Finally, the last justification for our choice is also given by Versteegh, when he mentions the following:

The topic of linguistic variation is restricted almost exclusively to writings such as these, as well as the books of the Arab historians, geographers and travellers. They sometimes inform us about different pronunciations in various areas and about the lexical variation in the areas which they visited.<sup>32</sup>

Exceptionally, we have also included in this research the analysis of all the *ḥattā* in the Quran: the religious book, although does not belong to the type of text previously mentioned, is a key element for the

<sup>28</sup> Except for the lyrical texts, also found continuously over time.

<sup>29</sup> This corpus is the Historical Arabic Dictionary Corpus, developed by Al-Said in 2010, as we will explain in section 3.2.

<sup>30</sup> Versteegh, “The Arabic Language”, p. 70.

<sup>31</sup> Versteegh, “The Arabic Language”, p. 70.

<sup>32</sup> Versteegh, “The Arabic Language”, p. 130.

formation and fixation of the standard Arabic, and it will be the temporal point of departure in our diachronic approach.

(ii) *Time*. In our study we have analysed data from six historical periods of approximately one century each. These periods represent key phases in the evolution of the Arabic historiography as well as in the history of the Arab and Muslim world, from the 7th to the 20th century. As we lack a periodization of the Arabic language, we have considered other extralinguistic elements for the selection of these centuries; to that end, the development and rise of the chosen types of text (history and geography books, genealogies and travel literature) have been taken into account, as it has also been the cultural splendour of the different areas of the arabophone world.

a) *7th century. The Quran*. Our diachronic point of departure lies in the Quran, because it represents, together with the pre-Islamic literature, the foundational text for the fixation and standardization of Classical Arabic<sup>33</sup>. We do not consider the Quran as the starting point of the Arabic language, but as an extremely important example of a specific type of language with a literary and elevated register that, needless to say, does not account for the variety of spoken languages of the time. In the Quran we will not find all the morphosyntactic and semantic possibilities of *ḥattā*; however, all forms and structures present in the religious book are susceptible to be repeated over the centuries.

b) *9th century. Beginnings of the Abbasid Caliphate*. Under the Abbasids, the Golden Age of Islam started, and the Arabic became the language of the Empire and the culture. During this period, there was a development of the Arts and the Sciences, and our selected books come from the centre of the Abbasid Caliphate in that moment: the modern territories of Iraq, Iran and Arabia.

c) *13th century. End of the Abbasid Caliphate*. In 1206, with the Mongol invasion, it starts the decline of the splendid Abbasid dynasty. It is important to analyse texts not only from the beginning, but also from the end of this golden period: the Abbasid Caliphate spans four centuries in which a stunning written production flourishes, a great number of translations that promote the inclusion of new words and

<sup>33</sup> Ferrando, “Introducción a la historia”, p. 82.

structures are carried out, and Arabic becomes a means for education.<sup>34</sup> For this historical period, texts from the opposite ends of the Muslim territory have been selected: Iraq and al-Andalus.

d) *14th century. Marinid dynasty and Mamluk Sultanate of Cairo.* In the Maghreb, the Berber Marinid dynasty was ruling and the travel journals proliferated. On the other hand, during the Mamluk Sultanate of Egypt (1250-1517) there was a boom of historiography, as the Mamluk kings were very interested in the production of chronicles. Finally, Ibn Kathir, an author that is also included in our study, was the last great member of the Syrian historiographical school of the *‘ulamā*. In this period we have considered, therefore, texts from al-Andalus, the Maghreb, the Levant and Egypt; these were areas in which the historical accounts were still flourishing in spite of the general decline of the Arabic literature.

e) *19th century. Renaissance.* Under the Ottoman Empire, especially from the middle of the 14th century, the production of Arabic historiography widely decreases, as the Turkish language becomes the language for the writings in the majority of fields. We will have to wait until the Napoleonic invasion of Egypt (1798-1801) to see a revival of the national conscience, and to witness a textual production of great interest for the study of the Arabic language. It is in this century when politics for the learning of Arabic by the population are developed, and when we find an important increase of the publication of Arab texts. For this reason, we propose this period as our fifth century for study, with texts from Iraq, Arabia and Egypt.

f) *20th century. Contemporary period.* It is important to include contemporary texts when dealing with language change, and therefore, the 20th century will be our last period of study. It is in this moment when an important standardization of Arabic is carried out, especially through the media, but also with the creation of language academies and the promotion of the access to education. We also find influence from European languages (especially English), not only on a lexical level, but also phonological, morphological and syntactic. In this sense, we align with Melis, Fores and Bogard<sup>35</sup> and with Pons Bordería<sup>36</sup>, who propose the

<sup>34</sup> Chejne, "The Arabic Language", pp. 64-80.

<sup>35</sup> Melis, Flores, and Bogard, "La historia del español".

<sup>36</sup> Pons Bordería, "El siglo xx como diacronía".

20th century as a relevant historical period for the analysis of diachronic evolution. We have considered texts from Iraq, the Levant and Egypt.

(iii) *Origin of the author*. The third of the macrovariables controlled in this analysis has been the geographical origin of the different authors. We have adopted a pan-Arabist and pan-Islamist perspective and, therefore, works from all areas of the Arabophone world have been included. *Fuṣḥā* is considered quite a homogeneous language, independently of the geographical area where a specific text comes from. We will not reject beforehand this assumption, but we believe it is convenient to register the origin of every text, especially on a historical analysis. Indeed, we will observe throughout the study that some linguistic innovations are very linked to concrete geographical areas.

(iv) *Type of language*. In this study we only consider texts in Classical and Modern Standard Arabic (that is, in *fuṣḥā*) throughout history. We defend the need to integrate the study of Arabic dialects for a comprehensive analysis of language change: dialects are key elements to better understand the history of this Semitic language, as they can offer information about previous stages of the language<sup>37</sup> and complement what the written texts tell us. However, for space constraints, here we will only focus on the analysis of *fuṣḥā*.<sup>38</sup>

### 3.2. *Corpus and selection of examples*

Once established the main characteristics of the sources where we will study the element *ḥattā*, we undertook the selection of the corpus, which should allow us selecting the analysis texts based on these previously determined macrovariables. Historical corpora in Arabic, in comparison with those existing for many other languages, are very limited. Ideally, we would need a diachronic corpus that includes texts from all centuries, continuously, from the most ancient periods until now. After examining the historical corpora in Arabic at our disposal (ArabiCorpus; King Abdulaziz City for Science and Technology Arabic Corpus [KACSTAC]; King Saud University Corpus of Classical Ara-

<sup>37</sup> Owens, “A linguistic history”, pp. 8-13.

<sup>38</sup> A comprehensive study of *ḥattā* in the Arabic dialects can be found in Medea-García, “Gramaticalización y cambio lingüístico”, pp. 241-302.

bic [KSUCCA] and Historical Arabic Dictionary Corpus [HADC]), we reached the conclusion that only the HADC met the requirements for our study. The Historical Arabic Dictionary Corpus (HADC) was developed by Almoataz B. Al-Said at Cairo University in 2010, and it is still unpublished. It is a historical corpus of Classical Arabic with more than 16 million words, and it includes texts from the 2nd until the 20th century, continuously.<sup>39</sup> It can be accessed exclusively contacting its author, and it is the only corpus in Arabic that allows selecting texts by period, century, author, type of text, geographical area and number of words.

Seven different aspects characterize the methodology for the selection of tokens—in this case, of examples with the particle *ḥattā*—. (i) We have worked with a representative sample, that is, big enough to provide significant results: concretely, we have analysed at least 100 examples of *ḥattā* for each century, with a total of 731 tokens. (ii) We have also built an equilibrated working corpus, with approximately the same number of examples of *ḥattā* per century and author: around 30 to 40. (iii) We have selected tokens of the particle throughout all the text, not exclusively from the beginning or the end, in order to avoid possible introductions or conclusions with a formulaic or specific nature.<sup>40</sup> (iv) We have created a working corpus with texts from different authors, to guard, to the extent possible, against idiolectal bias. (v) It is a working corpus that contains texts belonging to different genres within the broad selected type of text: that is, in each historical period, we have tried to include historical and geographical texts, as well as genealogies and travel literature. (vi) Finally, authors from different geographical origins throughout the Arabophone world have been selected, and (vii) all lyrical passages found have been dismissed. The following table gathers the basic information about the texts used for the study, as well as the number of *ḥattā* that can be found in every book and that have been analysed for this research.

<sup>39</sup> For a more detailed description of its characteristics and the methodology to consult the corpus, see Medea-García, “Gramaticalización y cambio lingüístico”, pp. 53-70.

<sup>40</sup> We have created a formula to guarantee the random selection of all tokens: “total # of *ḥattā* in the book / total # of *ḥattā* needed = sampling period (distance of apparition between the selected *ḥattā*)”.

Table 1: Texts selected for the study

Century	Book	Year	Type of text	Region	Analyzed <i>ḥattā</i>	Total # of <i>ḥattā</i>
7th	<i>al-Qurʾān al-Karīm</i>	738	Religious text	Arabia	142	142
9th	<i>Ġamḥarat ʾānsāb al-ʿarab</i>	820	Genealogy	Iraq	7	13
	<i>Futūḥ al-Shām</i>	823	History	Iraq	71	1022
	<i>Nasab Quraiṣ</i>	851	Genealogy	Arabia	21	197
	<i>al-Masālik wa-al-mamālik</i>	894	Geography	Iran	71	77
13th	<i>Riḥlat Ibn Ḡubair</i>	1217	Travel literature	Al-Andalus	47	116
	<i>Muʿjam al-buldān</i>	1229	Geography and History	Iraq	16	2041
	<i>al-Kāmil fī al-tārīḥ</i>	1233	History	Iraq	37	4835
14th	<i>al-ʿibr wa-dīwān al-mubtadā wa-al-ḥabr</i>	1360	History	Maghreb	25	3106
	<i>al-Bidāya wa-al-nihāya</i>	1373	History	The Levant	40	7581
	<i>Nuḡdat al-ḡurāb fī ʾulālat al-iḡtirāb</i>	1374	Travel literature	Al-Andalus	23	74
	<i>Riḥlat Ibn Baṭṭāḥ</i>	1377	Travel literature	Maghreb	18	307
19th	<i>ʿaḡāʾib al-aṭār fī al-tarāḡim wa-al-aḥbār</i>	1824	Genealogy and History	Egipt	42	1719
	<i>al-Dirar al-mufāḥir fī aḥbār al-ʿarab al-awāḥir</i>	1830	Genealogy	Arabia	17	21
	<i>Ġarāʾib al-iḡtirāb wa-nuḡat al-albāb</i>	1854	Travel literature and Geography	Iraq	41	155
20th	<i>Ḥalfiyat al-buṣār fī tārīḥ al-qirn al-tālī ʾaṣar</i>	1905	History	The Levant	40	583
	<i>al-Maṣṣil fī tārīḥ al-ʿarab qibal al-Islām</i>	1986	History	Iraq	51	3359
	<i>Tārīḥ al-ʿarab wa-ḥaḍarātihim fī al-Āndalus</i>	2000	History	Egypt	22	241

### 3.3. Analysis datasheet and relational database

In order to gather in a systematic and organized way all the information needed for the research, we have created a relational database (with the software *FileMaker Pro 12*) and an analysis datasheet<sup>41</sup>. This detailed datasheet has been designed to obtain all the morphosyntactic, semantic and pragmatic information needed from each of the examples of *ḥattā*. In total, there are 30 variables taken into account for every token of *ḥattā*: the lexical aspect of the verb preceding *ḥattā*; the grammatical category of *ḥattā* according not only to the Western tradition, but also to the Arabic Grammatical Tradition; the existence of a negation preceding or following *ḥattā* or the fact of being an example in direct or indirect speech, among many others.<sup>42</sup>

<sup>41</sup> This database is available in the following *GitHub* repository: <https://github.com/lucia-medea/hatta-database>.

<sup>42</sup> For a complete description and explanation of all the linguistic variables taken into account, see Medea-García, “Gramaticalización y cambio lingüístico”, pp. 71-90.



### 3.4. Recapitulation

With this methodology, we have analysed 731 examples of *ḥattā* in *fuṣḥā* spanning 1400 years, using one of the most complete historical corpora of the Arabic language, the HADC. The study considered 30 different variables that have been organized in a relational database. We would like to mention, therefore, some of the most relevant aspects of this methodological proposal.

(i) We have used and adapted<sup>43</sup> the HADC, probably one of the most exhaustive, suitable and also unknown historical corpora of Arabic nowadays, and the only one that allows having so many variables controlled in an automatic or semi-automatic way.

(ii) The analysis includes texts from historical periods barely taken into account in other diachronic approaches to this Semitic language, such as the 13th century, particularly important to see the evolution of our targeted element during the flourishing Abbasid Caliphate, or the 14th century, essential when dealing with historiography and travel literature texts.

(iii) A great number of variables, not considered in previous studies, have been taken into account in our approach, such as the type of text, the geographical origin of the author or the specification of historical periods to include. The analysis provided is, therefore, more detailed and rigorous.

(iv) This methodology contemplates a qualitative but also a quantitative analysis, thanks to the resources that computers and statistics offer. The combination of methodologies and tools from corpus and computational linguistics strengthens the results and conclusions reached.

(v) Although we work mainly with theoretical frameworks and tools developed in the Western world, we take into account the Arabic Grammatical Tradition when dealing with the understanding and evolution of our particle. The AGT is one of the most ancient and prolific grammatical traditions, and it is necessary to understand what has been said about a specific element if we want to provide a comprehensive and exhaustive account of it.

(vi) Finally, our analysis has been data-driven: we have collected, observed and understood the linguistic data, and only after that we have

<sup>43</sup> Medea-García and Al-Said, “The HADC and its Suitability”.



selected the kind of theoretical frameworks that can help us explain, in a deeper way, all the information found.

In conclusion, we consider this methodological proposal is solid and rigorous. The selected macrovariables allowed us to generate a controlled selection of texts to analyse; the HADC is, in our opinion, the most appropriate corpus to develop a diachronic study of Classical Arabic; the sample of examples with *hattā* is representative, equilibrated and enough to provide a first approach. The 30 linguistic variables inserted in the datasheet guarantee a detailed and comprehensive analysis of this element, and, finally, the creation of the relational database to gather all this information facilitates the searching tasks as well as the exploration of linguistic evolutionary patterns, among many others.

#### 4. Grammaticalization(s) of *hattā* in Classical Arabic

In this section, we will briefly describe and analyse the evolution of the semantic values of *hattā* throughout the centuries; concretely, we will study the behaviour of this particle in the six periods of study previously mentioned (7th, 9th, 13th, 14th, 19th and 20th centuries) and in the books about history, geography, genealogy and travel literature. Moreover, we will review the examples extracted from the corpus in the light of the grammaticalization studies to have a deeper understanding of the historical pathway of this element.

##### 4.1. Semantic values of *hattā*

We will start by offering a semantic classification of the kinds of *hattā* found in the corpus: we will name them, translate them into English, and mention the part of speech they belong to. Finally, we will provide an example from our corpus. The semantic criterion has been essential to classify and understand the variety of senses and uses that this particle presents. In total, there are five differentiated semantic values of *hattā* in the Classical Arabic texts throughout the 14 centuries contemplated in our analysis (7th - 20th).

(i) *Temporal*. It is translated as ‘until’, and it codifies the values of temporal trajectory and limit. It can be a preposition and also a temporal subordinating conjunction:

(1) *faḍammahu 'ābūhu al-ḥusayn 'īlayhi ḥattā mā*

‘And his father Hussein hugged him until he died’ (13, 9th century)<sup>44</sup>

(i') *Spatial*. It is also translated as ‘until’ and it codifies the values of spatial limit and trajectory. Similarly, it can be a preposition and a spatial subordinating conjunction. As it is nearly always linked to the temporal *ḥattā*, we include these two first values in the concept of “spatiotemporal”.

(2) *wa-l-minṭaqa al-ḡanūbiyya al-mumtadda min ḡanūb Ṣan'ā' ḥattā madīna 'adan*

‘And the south region that spans from the south of Sana'a until the city of Aden’ (525, 20th century)

(ii) *Purpose*. It is translated as ‘in order to’, ‘so that’, and it is a purpose subordinating conjunction. It mainly codifies the concept of trajectory, while the idea of limit remains in a vague way (as it is not sure that the limit will be reached).

(3) *tumma yaqfulūn fayuqīmūn 'ilā ḥamsa wa- 'iṣrūn yawman wa-hiya baqiyyat ḥuzayrān wa-ḥamsa min tammūz ḥattā yaqwā wa-yasman al-zahr*

‘Then they come back and stay for 25 days, that is, the rest of the month of June and 5 from the month of July, so that [the horse] gets stronger and gains weight on its back’ (88, 9th century)

(iii) *Consecutive-intensifier*. It is a consecutive conjunction with an intensifier value, that could be translated as ‘to the point that’, ‘so much that’.

(4) *wa-qad iṣṭhara Daḡfal fī n-nasab, ḥattā ḍuriba bihi l-maṭal fī l-naṭr wa-fī ṣ-ṣi'r bi-sa'at 'ilmihī bih*

‘And Daḡfal was known [for his wisdom] in genealogy, to the point that he was mentioned as an example to follow in writing prose and poetry thanks to his broad knowledge on it’ (552, 20th century)

<sup>44</sup> The number that appears after the translation refers to the identificatory number of this example in the relational database. It is, therefore, easy and quick to find the concrete example to have all the information about the author, book, century, origin, type of text, etc. The relational database, as mentioned before, can be found in the following *GitHub* repository: <https://github.com/lucia-medeia/hatta-database>.

(iv) *Scalar*. This kind of *hattā* is a scalar adverb with the sense of ‘even’ (or ‘not even’, depending on the context) and, therefore, it codifies the limit of a scale.

(5) *wa-yaḥṣāhu ḥattā n-nisāʾ fī l-buyūt wa-huwa faʾit*

‘And even women inside their houses feared him [while he was passing by]’  
(331, 19th century)

(v) *Temporal-disruptive*. It is translated as ‘when (suddenly)’, and it works as a relative temporal adverb. It codifies the limit of an action A that cannot be finished, and at the same time it indicates the origin of an action B—that precisely prevents the action A to be accomplished—. Therefore, this kind of *hattā* marks the limit or end of the action preceding the particle as well as the beginning of a new action.

(6) *wa-lam yakad yaqirr al-qarār, wa-lā tunza ʿal-ḥifāf, ḥattā ġamara min aṭ-ṭaʾām al-baḥr*

‘As soon as we arrived, and the shoes had not yet been taken off, when [suddenly] the sea submerged the food’ (272, 14th century)

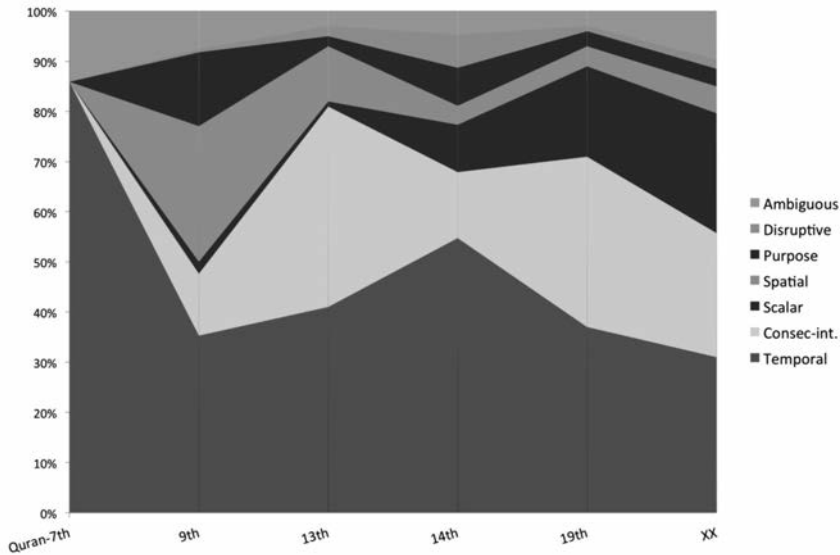
In our analysis we also include the “ambiguous” *hattā*. It is not a specific kind of *hattā*, but it includes the examples that have two perfectly possible readings, in which the context is not enough to distinguish the sense meant by the author. They represent a 6% of our corpus, and they are of great linguistic interest for a diachronic approach, because ambiguity can be a trigger of language change. We provide an example of ambiguity between the temporal (‘until’) and the purpose (‘so that’) values:

(7) *faʾdan lī fī ittībā ʾihim ḥattā ʾaruddahum ʾalayk*

‘Then allow me to follow them *in order to* / *until* I bring them back again’  
(199, 13th century)

After analysing 731 cases of *hattā* from the 7th to the 20th century, we present here a summarized graphic with the percentage of apparition and the evolution of the semantic values of this particle.

Image 2: Semantic values of ḥattā



#### 4.1.1. The Quran - 7th century

Our historical analysis begins in the 7th century with the Quran, and three main ideas arise from the study of ḥattā in this period.

(i) The one and only fixed value of the particle is the temporal ‘until’, which also represents the principal use, as it includes the 86% of the examples in the religious book.

(ii) All the rest of examples present ambiguous readings, with two possible interpretations and where a temporal value (‘until’) is always possible. Concretely, we find examples of ambiguity between the temporal (which, as already stated, is always possible) and the purpose value, the exceptive-conditional value<sup>45</sup> and the consecutive-intensifier

<sup>45</sup> The “exceptive” value of ḥattā can be translated by ‘unless’, and it combines the notion of condition and exception. It works as a conditional-exceptive subordinating conjunction. In our full corpus there is not a single example of this kind of ḥattā with a univocal reading; only 7 out of the 731 tokens analysed have an ambiguous sense, where both the temporal (‘until’) and the conditional-exceptive (‘unless’) values are possible. For this reason, we have not considered this value as a separated kind of ḥattā, even if old and contemporary grammars include it as a kind of ḥattā ḡarra (ḥattā al-istiṭnā). See Sadan, “The subjunctive mood”, p. 217.

value. These numbers show that the temporal *ḥattā* ‘until’ was a very stable and consolidated value, and at the same time, that different language variations were happening. Although all translations from the corpus examples have been carried out by the author of this article, in the case of the Quran we have used and compared the seven most accepted translations by the international community<sup>46</sup>, offered in the Quranic Arabic Corpus electronic resource. Here we provide an example of an ambiguous reading between the temporal and purpose values:

(8) *wa-lā yazālūna yuqātilūnakum ḥattā yaruddūkum ‘an dīnikum*

*Sahih*: ‘And they will continue to fight you until they turn you back from your religion if they are able’

Muhammad Sarwar: ‘(The pagans) still try to fight you to make you give up your religion’ (374, 2:217, Quran)<sup>47</sup>

(iii) In third place, we find an interesting example that is, in our opinion, the germ of the scalar adverb ‘even’. The AGT defends that in the Quran there is not any case of this kind of *ḥattā*, designed as *ḥattā ‘āṭifa*<sup>48</sup>. However, there is an example in the religious book where *ḥattā* is fluctuating between a consecutive-intensifier and, surprisingly, a scalar value; the above-mentioned reference translations confirm it:

(9) *wa-’in yaraw kulla āyatin lā yu’minū bihā ḥattā ‘idā ḡā’ūka yuḡādilūnaka yaqūlu l-laḏīna kafarū ‘in haḏā ‘illā ‘asāṭīru l-’awwalīna*

*Sahih*: ‘And if they should see every sign, they will not believe in it. Even when they come to you arguing with you, those who disbelieve say, “This is not but legends of the former peoples”’

*Yusuf Ali*: ‘if they saw every one of the signs, not they will believe in them; in so much that when they come to thee, they (but) dispute with thee; the Unbelievers say: “These are nothing but tales of the ancients”’ (395, 6:25, Quran)

<sup>46</sup> The translations are provided by the Quranic Arabic Corpus, an annotated linguistic resource developed by the Language Research Group of Leeds University. The seven used translations of the Quran are the following: *Sahih International*, “The Quran”; *Ali*, “The Holy Quran”; *Shakir*, “The Holy Quran”; *Sarwar*, “The Holy Qur’an”; *Khan*, “The Noble Quran”; *Pickthall*, “The Meaning of the Glorious” and *Arberry*, “The Koran Interpreted”.

<sup>47</sup> When examples from the Quran are offered, at the end we indicate, in brackets, the reference of the number in the database; then, the number of the surah, colon, and the number of the ayah.

<sup>48</sup> See *Ibn Hišām*, “*Muḡnī al-labīb*”, p. 146, among others.

This first period of analysis shows that *ḥattā* is already immersed in a grammaticalization process. More precisely, we can observe a semantic bleaching of the particle, especially if we take into account the proposed etymological origin, the noun *ḥadd* ‘limit’, ‘border’. This lexical referential meaning of a noun is transformed into a subordinating conjunction and preposition that codifies the spatiotemporal concepts of trajectory and limit (‘until’). This semantic bleaching will continue in the following centuries, when *ḥattā* adopts more abstract and more cognitively complex values. The tokens of *ḥattā* in the Quran also show a favourable scenario for language change, as there is a significant percentage of ambiguous cases.

#### 4.1.2. 9th century

We observe, in the 9th century, an important decrease in the temporal *ḥattā* (‘until’), from 85% in the Quran to 35%, and also in the ambiguous *ḥattā* (from 14% to 6,5%). At the same time, there is an increase of the variety of values codified by this element: we find two completely new senses, as well as three other values that become established, already with univocal readings. Generally speaking, we can make three main observations.

(i) Two new values of *ḥattā* come to light in our corpus from this period. Firstly, the spatial *ḥattā* appears vigorously, as it goes from a 0% of cases to a 27%. More than 70% of these spatial *ḥattā* come from geography texts, which points to the weight of the type of text in the development of the semantic values of the particle.

A completely new value also seems to appear in this century: the temporal-disruptive *ḥattā* (‘when [suddenly]’). It looks like an example of reanalysis, because the first —and only the first— examples observed in the corpus have an ambiguous reading divided between the temporal and the disruptive sense. This kind of cases could be considered “bridging contexts”<sup>49</sup>:

(10) *qāl: famā marrat al-’ayyām ḥattā qadima ’anas raḍiya l-lāh  
’anhu yubašširuhu biqudūm ’āhl al-yaman*

‘He said: ‘it did not take long *until* / *when* Anas arrived and informed him that the people from Yemen were coming’ (111, 9th century)

<sup>49</sup> Evans and Wilkins, “In the mind’s ear”, p. 5 and Heine, “On the role of context”, p. 84.

This possible double interpretation may have motivated the receiver to choose one of them and, when using this structure again, may have generated an unambiguous syntactic context:

- (11) *qāl famā statamma kalāmuh ḥattā badat al-ḥayl ka 'annahā s-sayl*

‘And he had not finished talking yet when [suddenly] the herd of horses appeared as if they were a torrent’ (127, 9th century)

This example, from the text *Futūḥ al-Šām*, shows the specific contexts in which the univocal readings of the disruptive *ḥattā* appear, which present these characteristics: (i) there is always a negation preceding *ḥattā*; (ii) there is always a positive sentence following the particle; (iii); the aspect of the verb preceding *ḥattā* is perfective; and, finally, (iv) the lexical aspect of the verb preceding *ḥattā* is always achievement or accomplishment.

This use of *ḥattā* is especially interesting, as it seems an example of pragmaticalization<sup>50</sup>: a metaphorical-metonymical process in which the speaker manipulates the forms after its use in specific discursive contexts. The pragmatic inferences become codified and conventionalized. Indeed, in the following centuries we will not find any more examples of disruptive *ḥattā* with a double meaning<sup>51</sup>.

(ii) Secondly, from this 9th century we must highlight that two other kinds of *ḥattā* seem to experience a process of subjectivization. Subjectivization is considered a process of a unidirectional nature, a tendency to go “from meanings grounded in more or less objectively identifiable extralinguistic situations (...) to meanings grounded in the speaker’s attitude to or belief about what is said”<sup>52</sup>.

On the one hand, the consecutive-intensifier *ḥattā* represents a first step towards this process of subjectivization. Consider the following example from the book *Ġamḥarat ansāb al-‘arab*:

- (12) *waqāl: “lā ašba ‘allāh baṭnak, ‘amā takfika l-ḥilāfa ḥattā taḡī’ fa-taṭlub ad-dār?”*

‘And he said: “Is it not enough for you the power, to the point that you come here and [even] ask for the house?” (105, 9th century)

<sup>50</sup> Diewald, “A model for relevant types”; Diewald, “Grammaticalization and pragmaticalization” and Diewald, “Pragmaticalization (defined) as grammaticalization”.

<sup>51</sup> Except for a case in the 20th century.

<sup>52</sup> Traugott and König, “The semantics-pragmatics of grammaticalization”, p. 189.

In this case, *ḥattā* is used both to intensify the preceding clause and to introduce a consequence of this intensification ('it is not enough, to the point that you even ask...'). It is, therefore, a case of discourse subjectivization, in which the author highlights the importance of some aspect; in this example, the greed of the receiver of the message.

On the other hand, the first univocal example of *ḥattā* as a scalar adverb, which in our corpus appears in the 9th century, also represents a process of subjectivization. The example comes, as the previous one, from the text *Ġamḥarat ansāb al-ʿarab*:

- (13) *fa-ḡaḍiba wulduhu fa-ḥaraḡū faḍarabūhum ḥattā an-nisāʾ faḍlan ʿalā r-riḡāl*

'His children got angry, went out and beat them, even the women, besides the men' (106, 9th century)

We can observe, thanks to the comment "besides the men", that it is an incipient example, as it still needs a pleonastic explanation. However, it shows the high degree of the author's implication, who considers the action of beating women as the lowest in the scale. As subjectivization has a unidirectional tendency<sup>53</sup>, we can deduce that values of *ḥattā* such as the consecutive-intensifier or the scalar arose later than the temporal one.

(iii) The third aspect worth mentioning about this 9th century is that two of the values that appeared only with an ambiguous reading in the Quran consolidate in this period, as they increase in frequency and are used in univocal contexts. It is the case of the purpose and consecutive-intensifier values, which in the previous period could always be understood as temporal. The semantic bleaching mentioned in the preceding period continues and, therefore, the particle adopts even more grammatical, more abstract and more complex values.

#### 4.1.3. 13th century

In the 13th century, that marks the end of the Abbasid Caliphate, there are two aspects that deserve attention.

(i) The first one is the remarkable rise of the consecutive-intensifier *ḥattā*: from a 12% in the 9th century, to a 40% in this period. It is interesting to observe that more than the 72% of these examples come

<sup>53</sup> Traugott, "(Inter)subjectivity and (inter)subjectification", p. 32.



from travel literature texts from al-Andalus. We provide an example from the book *Rihla Ibn Ġubaīr*:

(14) *wa-sakana l-baħr ħattā ħuyyila li-nāziriħ šaħn zuġāġ azraq*

‘And the sea calmed down to the point that it was imagined by the observer as a plain recipient of blue glass’ (229, 13th century)

It is precisely in this century, in the travel literature genre, and in texts from al-Andalus, where the consecutive-intensifier *ħattā* is developed and well-established. Indeed, in the travel literature books we find much more modalization, much more implication of the author, than in the rest of considered types of texts.

This kind of examples represents another step in the process of subjectivization; they are also a proof that the genre and the geographical origin are relevant factors in the development of the different semantic values of *ħattā*.

(ii) The second of the remarkable aspects is the nearly total absence of *ħattā* with an ambiguous reading: they only represent a 2% of all the cases, and they offer exclusively a temporal or purpose reading. This is the century with less amount and fewer kinds of ambiguous *ħattā*: it seems the Golden Age helped in fixing and establishing the different values of the particle.

#### 4.1.4. 14th century

Three main aspects need to be highlighted in this fourth period of study.

(i) Firstly, there is a remarkable increase of the scalar *ħattā*: from 1% in the previous century, to 10% here. The most interesting part is that 26% of the tokens of *ħattā* coming from Andalusian travel literature texts are scalar, while in other types of texts, the percentage of scalar *ħattā* is less than 5%. The great majority of these *ħattā*, therefore, developed in al-Andalus and in travel literature texts. It is worth noting that, nowadays, the scalar value is the most common use of this particle in nearly all Arabic dialects<sup>54</sup>.

(ii) On the other hand, the temporal-disruptive *ħattā* (‘when [suddenly]’) reaches, in this century, its maximum frequency: nearly a 7%. Again, if we consider exclusively the Andalusian travel literature texts,

<sup>54</sup> Medea-García, “Grammaticalización y cambio lingüístico”, pp. 241-302.

we observe that a 22% of *ḥattā* have this disruptive value; in the other type of texts, only a 2.4% of examples of this particle exhibit this kind of value.

(iii) Thirdly, we find in this period, for the first time in our entire work corpus, a prepositional *ḥattā* with a spatial value, which belongs to the Andalusian travel literature text *Nuḥḍat al-Ġurāb fī ‘ulālat al-igtirāb*:

(15) *fa-ṭṭarada bāb aš-šuhḥ ḥissan wa-ma ‘nan, wa-muwahḥadan wa-muṭannan ḥattā dukkaāla*

‘And the matter of the physical and psychological greed, individual and shared, continued until Dukkala [a village in Morocco]’ (269, 14th century)

This value and this part of speech of *ḥattā* (that is, a spatial preposition), should represent, theoretically speaking, one of the first steps of *ḥattā* towards its process of grammaticalization. The lack of this kind of examples could be explained with two arguments. On the one hand, *ḥattā* already seemed inserted in a quite advanced process of grammaticalization in the Quran, where we find temporal but not a single spatial example of *ḥattā*. On the other hand, the preposition *‘ilā* seems to be the unmarked element used to codify, since ancient times, the spatiotemporal limit when a preposition is needed.

It is interesting to note that al-Andalus and travel literature texts are, therefore, a *locus* of language change and innovation: the consecutive-intensifier, scalar and disruptive *ḥattā* developed in those domains. And these kinds of *ḥattā* arose, precisely, through processes of subjectivization and pragmaticalization.

#### 4.1.5. 19th century

In the 19th century there are three main values of *ḥattā* that stand out: the temporal (37%), the consecutive-intensifier (34%) and the scalar (18%). The rest of the values present a marginal frequency, and the ambiguous readings of *ḥattā* are unusual, with only a 3% of the cases.

There is a remarkable aspect that shows how the grammaticalization process of *ḥattā* continues its path: at least two examples of the consecutive-intensifier value have lost the “intensifier” nuance, and only keep the consecutive value. We show a case from the text *al-Darar al-mufāḥir fī aḥbār al-‘arab al-awāḥir*:

- (16) *ḥakama muḥammad ibn sa 'ūd ḥattā 'aqāma 'alā taḥtihi muda, wa-ba 'd al- 'azīz ibnuh*

‘Muhammad Ben Saud was ruling and so he stayed in power for a while, and then his son Al-Aziz’ (858, 19th century)

We can observe that the consecutive-intensifier *ḥattā* looses the value of ‘limit’ —its etymological and most lexical value— and also the intensive nuance, while at the same time the consecutive value is accentuated (‘so’). This last value is more grammatical and more abstract, which indicates a step forward in the grammaticalization process of this particle.

#### 4.1.6. 20th century

The 20th century is a relevant period of study, even in a diachronic approach, as there are many elements that influence the language.

At first glance, the semantic values of *ḥattā* in this period seem similar to the ones in the previous century. Indeed, we find the same three main values —temporal, consecutive-intensifier and scalar— with about a 25% of frequency for each one of them. However, there is an element with a significant relevance: the cases of *ḥattā* with an ambiguous reading experience a remarkable increase, not only in its frequency, but also in the kind of double readings. We find now nearly an 11% of ambiguous *ḥattā*, while in the previous century there were only a 3%; we also find four different kinds of ambiguous readings, always fluctuating between a temporal and another value. What is more, two of these four kinds have only been found in the two first periods of study (7th and 9th centuries), showing that there has been a return of old constructions and expressions. Concretely, the *ḥattā* with a temporal or disruptive value only appears in the 9th century<sup>55</sup>, and the *ḥattā* oscillating between a temporal and a conditional-exceptive value only appears in the Quran (as well as in this 20th century), as it shows the following example:

- (17) *fahādīhi l- 'ibāra naṣṣ ṣarīḥ fī 'annahu lā yastaqīm 'amr al-muslimīn ḥattā yakūnū kamā kāna ṣ-ṣaḥāba raḍīya al-lāh 'anhum*

‘And this quotation [said by Abu Bakr] is a direct and honest text about the fact that the actions of Muslims are not right *unless/until* they are as the actions of the first Muslims, God bless them’ (665, 20th century)

<sup>55</sup> See example 10.

In this period we also observe, for the first time, examples of *ḥattā* with an undetermined value, that seem to need from other elements to have a clear meaning. We illustrate this case with a sentence from the book *Tārīḥ al-‘arab wa-ḥaḍarātihim fī al-‘andalus*, where *ḥattā* seems to have a temporal-disruptive value:

- (18) *mā ‘in waṣalū ‘ilā ‘Urš (Orch) fī ‘adnā kūra raya, ḥattā tarakū ‘aḥadahm mā ‘*

*l-hadāyā*

‘As soon as they arrived to Orch, the closest area in the region of Rayya, [it was] when they left one of them with the presents’ (783, 20th century)

This sentence needs from a copulative verb (*was*) in order to have a clear meaning. Furthermore, the temporal-disruptive *ḥattā* does not exhibit its characteristic “surprise” or “irruption” value, because it is now given by the previous construction *mā ‘in* (‘once’, ‘as soon as’). At the same time, we observe that this *ḥattā* adopts a more sequential value, and it looses a bit of its referential meaning: in fact, if we exclude the particle, the sentence is grammatical and, besides, its meaning remains the same (‘as soon as they arrived to x, they left...’), an aspect that proves the proximity of this *ḥattā* to a copulative nature.

The rise of the ambiguous *ḥattā* in this 20th century can be attributed to several linguistic and extralinguistic factors: (a) the fact that Modern Standard Arabic is the mother tongue of no one and, therefore, as it is not a language used in the everyday routine, the ambiguous constructions have not been solved in a natural way. Another possible cause can be (b) the author’s intention to return to classic and “pure” structures: as they have not been used probably in centuries, they may lead to ambiguity, misunderstandings and even to reanalysis. Finally, (c) the more generalized access to education and to the use of *fushḥā* favours a larger text production; as a consequence, the linguistic variation increases, as well as the ambiguous readings.

One of the main remarks that can be highlighted after this journey following the steps of *ḥattā* is the fact that the principles of *layering*, *persistence* and *divergence* are much more numerous and long-lasting in Arabic. These three principles are part of the five principles established by Hopper in 1991<sup>56</sup> in order to detect an ongoing process of

<sup>56</sup> Hopper, “On some principles”.

grammaticalization. The *layering* appears when, “within a broad functional domain, new layers are continually emerging; in the process the older layers are not necessarily discarded, but may remain to coexist with and interact with new layers”<sup>57</sup>. In the case of Arabic this layering is especially marked. Indeed, what has been established by the Quran (or the AGT) is not only correct, but desirable. Therefore, even if some elements or constructions experience processes of language change and can work and mean differently, the values fixed by the Quran can always remain in use. *Ḥattā*, in fact, continues to preserve its initial meanings, and all of them coexist with the new values of the particle that arose throughout the centuries. Actually, in the 20th century we find examples of all and each of the values that *ḥattā* has acquired over the years.

The principle of *persistence* appears when the grammaticalized forms preserve some characteristics of the original lexical meaning or even the syntax<sup>58</sup>. This persistence is clearly observed in all the uses and values that *ḥattā* develops. Its primitive meaning is the noun *ḥadd* ‘border’, ‘limit’, and the reference and inclusion of this concept is found in all the grammaticalized uses of the particle, as we will see in the following section (5).

The third principle we want to highlight is *divergence*: “when a lexical form undergoes grammaticalization to a clitic or affix, the original form may remain as an autonomous lexical element and undergo the same changes as ordinary lexical items”<sup>59</sup>. This concept refers, therefore, to the notional and functional split that sometimes happens in processes of language change, when from a unique lexical item can be formed two elements, one with a grammatical value and another that keeps the lexical meaning of the etymon. In *ḥattā* we also find this division between the lexical etymon *ḥadd*, uninterruptedly used with its original meaning of ‘limit’, ‘border’, on the one hand, and on the other, the grammatical element *ḥattā*, that experiences grammaticalization and subjectivization processes.

<sup>57</sup> Hopper, “On some principles”, p. 22.

<sup>58</sup> Hopper, “On some principles” and Cuenca, “La gramaticalización”.

<sup>59</sup> Hopper, “On some principles”.

## 5. A cognitive approach

The branch of cognitive linguistics (CL) is an appropriate and fruitful theoretical model to analyse the evolution of *ḥattā*. Four main reasons justify the election of this approach:

(1) First of all, this theoretical framework gives a lot of importance to the semantic aspect.<sup>60</sup> If we understand that the main function of language is to communicate, then the different linguistic phenomena must also be explained through semantic considerations. In this case, the study of the semantic values of *ḥattā* has been a key aspect in the descriptive analysis, as it has helped to organise the wide range of different meanings that this particle presents.

(2) The Conceptual Metaphor Theory (CTM), a sub-discipline of cognitive linguistics, is very explanatory for the analysis of *ḥattā*. Lakoff and Johnson explained, in their foundational book *Metaphors we live by*, that “the essence of metaphor is understanding and experiencing one kind of thing in terms of another”<sup>61</sup>. One of the most interesting points is the directionality of metaphors: “specifically, we tend to structure the less concrete and inherently vaguer concepts (like those for emotions) in terms of more concrete concepts, which are better delineated in our experience”<sup>62</sup>. This theory is particularly useful to address issues such as the polysemy of *ḥattā*, because the directionality of the metaphorical process helps us outline the diachronic path of its meanings.

(3) In third place, CL also defends that language is based on usage. Our study is in line with this assumption; indeed, we have found and adapted a historical corpus of Arabic, and we have provided real examples of written language to ground our analysis.

(4) Finally, we have adopted a cognitive approach also because this framework rejects many of the traditionally accepted dichotomies in the field of linguistics. Therefore, semantics and pragmatics are not conceived as clearly differentiated areas; the linguistic knowledge is not understood without the encyclopaedic knowledge; and, last but not least, it breaks with the dichotomy between synchrony and diachrony.

<sup>60</sup> Maldonado, “La gramática cognitiva”, p. 213.

<sup>61</sup> Lakoff and Johnson, *Metaphors we live by*, p. 5.

<sup>62</sup> Lakoff and Johnson, *Metaphors we live by*, p. 112.

In this approach, however, we will also adopt proposals linked to other branches of linguistics, that often have been integrated in cognitive analysis. Specifically, we will include methodologies and proposals from (a) grammaticalization studies, (b) linguistic typology and (c) corpus linguistics.

(a) In relation with the grammaticalization studies, we understand grammaticalization as a mainly semantic process, following Heine and Kuteva<sup>63</sup>. We will also assume the unidirectionality hypothesis, that is, the idea that if a grammaticalization process occurs, it will be in the direction of lexical word to grammatical word (or grammatical to more grammatical), and not the other way around. Another of the proposals from this framework relevant to the current analysis is the existence of grammaticalization channels, the evolutionary pathways experienced by the elements inserted in a grammaticalization process. It is worth mentioning that some of these channels “are universal in the sense that development along them occurs independently in unrelated languages”<sup>64</sup>. Finally, we highlight, within the grammaticalization framework, the studies about subjectivization developed mainly by Traugott<sup>65</sup>: as she proves, language change can be triggered by the author’s implication in the linguistic form of the message. This idea is essential for the understanding of some semantic values of *ḥattā*.

(b) Secondly, we also include a typological approach in this cognitive analysis. We have mainly used Heine and Kuteva’s work: *World Lexicon of Grammaticalization* (WLG), where the authors try to gather “many data from as many different languages as possible and to avoid theoretical biases – as far as this is possible and feasible”<sup>66</sup>. The purpose of this typological contrast is twofold: on the one hand, if the evolutionary patterns proposed for *ḥattā* are found in other languages too, then it strengthens our analysis. On the other hand, and maybe even more important, we would like to remind that in the WLG “over 400 processes relating to the evolution of grammatical categories are dis-

<sup>63</sup> Heine and Kuteva, “World Lexicon”, p. 3.

<sup>64</sup> Bybee, “Cognitive processes”, p. 256.

<sup>65</sup> Traugott, “Meaning-change in the development”; Traugott, “On the rise of epistemic meanings”; Traugott, “From less to more situated”; Traugott, “Subjectification in grammaticalization” and Traugott, “Semantic change”.

<sup>66</sup> Heine and Kuteva, “World Lexicon”, p. 1.



cussed, using data from roughly 500 different languages”<sup>67</sup>. However, not a single example has been dedicated to Arabic<sup>68</sup>. The present approach provides historical data from a language barely studied from a diachronic point of view. And these data can contribute to the search and detection of universal patterns of language change, both by strengthening grammaticalization channels already found in other languages, or by proposing new evolutionary pathways.

(c) Thirdly, corpus linguistics is also used in this approach, as all the material to ground our proposal comes from real examples, extracted from the Historical Dictionary Arabic Corpus.

We believe that a combination of proposals from cognitive and functional linguistics, in which we include the typological contrast as well as the corpus linguistics, can offer a much more profound, coherent and rigorous explanation of the history of *ḥattā*, and can help us reach a better understanding of the evolutionary processes of languages.

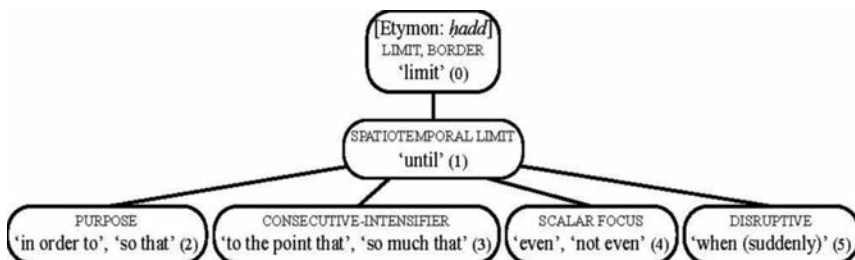
### 5.1. Conceptual map of the semantic values of *ḥattā*

We have developed a conceptual map with the data from our work corpus; a map that links and organizes the different semantic values developed by the particle *ḥattā* in Classical Arabic throughout fourteen centuries. The map can be divided in three stages, from top to bottom, that approximately correspond to their order of appearance and percentage of frequency in the language over the centuries. These three stages also account for the degree of grammaticalization acquired by the values of the particle *ḥattā*. As we descend in the map, values become more abstract, cognitively more complex and also more subjective and valorative.

<sup>67</sup> Heine and Kuteva, “World Lexicon”, p. 1.

<sup>68</sup> Indeed, in the whole *World Lexicon of Grammaticalization* we only find mentioned a couple of examples in Arabic, in which no evolutionary process at all is detailed, as it happens in page 41 or 275. It occurs something similar in the more recent *Oxford Handbook of Grammaticalization*, by Narrog and Heine (2011).



Image 3: Conceptual map of *hattā* in Classical Arabic

### 5.1.1. Stage I: etymological origin of *hattā*

As it has already been explained in Section 2, we consider that the closest origin of *hattā* would be linked to the Arabic noun *hadd* ‘limit’, and, at the same time, the etymological origin of both elements should be found in the pan-Semitic root ‘*d*’. This proposal would link Corriente, Pereira and Vicente’s hypotheses with Al-Jallad’s.

### 5.1.2. Stage II: spatiotemporal limit

The value of spatiotemporal limit is the first and main value of *hattā* found in our work corpus: it includes nearly all the Quranic *hattā*, and it is the predominant value throughout the centuries, although its trajectory is clearly descending. Example (1) illustrates this kind of *hattā*. We propose, alongside Al-Jallad, an evolutionary pathway from the noun ‘limit’ or ‘border’ (*hadd*) to the spatiotemporal preposition and conjunction with the value ‘until’:

(19) LIMIT > SPATIOTEMPORAL LIMIT

This pattern has been also found in other languages. In Heine and Kuteva’s words, the path BOUNDARY > UNTIL “appears to be another instance of a more general process whereby relational nouns give rise to relational (typically spatial or temporal) grammatical markers”<sup>69</sup>. The authors state that only examples from Africa have been found, such as in Swahili and Moré, both belonging to the Niger-Congo languages.

<sup>69</sup> Heine and Kuteva, “World Lexicon”, p. 75.

However, this pattern can also be found in other language families: in Italian (*fine* ‘limit’, ‘end’ > *fino* ‘until’), Catalan (*fine* ‘limit’, ‘end’ > *fins* ‘until’), or in Arabic, with the example of *ḥattā*.

We detach ourselves from the traditional proposal that links the spatiotemporal *ḥattā* with the supposed verbal root *ḥtt* ‘scrap off, rub off’<sup>70</sup>. Likewise, we do not agree with a more recent proposal by Esseesy, who defends that *ḥattā* “grammaticalizes along the cline indicating the pathway for semantic change for *ḥattā* predicted by Heine and Kuteva (2002,41): ALLATIVE > UNTIL”<sup>71</sup>. In fact, we have not found, neither in our work corpus nor in the consulted works, examples of *ḥattā* that could be included in the category ALLATIVE proposed by Heine and Kuteva<sup>72</sup>, equivalent to ‘to’, ‘direction towards’.

### 5.1.3. Stage III: rest of values

This last stage of the conceptual map is composed of the different values that *ḥattā* has been acquiring over the time; values that have developed after the concept of SPATIOTEMPORAL LIMIT. These values, numbered in the map (2-5), are more complex, more abstract and more valorative than the previous ones, in line with what the theory predicts.

(2) Purpose value. The second of the values developed by *ḥattā* is the purpose ‘in order to’. It points to the end, the objective or intention when you do something. Examples of spatiotemporal limit and purpose can be very close: “Claire studied *until* she became a teacher / Claire studied *in order to* become a teacher”. A main difference would be that purpose does not codify the arrival to the objective or goal, to the established ‘limit’, but it only projects it. The value of “spatiotemporal limit”, however, implies a continuous trajectory until that marked end. The evolutionary process would be the following:

(20) LIMIT > SPATIOTEMPORAL LIMIT > PURPOSE

It should be mentioned that in the *World Lexicon of Grammaticalization* the category LIMIT (or UNTIL) does not appear as a source element for the value of PURPOSE; in contrast, they offer eight different categories, such as ALLATIVE, BENEFACTIVE, GIVE, among others, as de-

<sup>70</sup> Lane, “An Arabic-English Lexicon”, p. 508b.

<sup>71</sup> Esseesy, “Grammaticalization of Arabic”, p. 319.

<sup>72</sup> Heine and Kuteva, “World Lexicon”, p. 17.

rived from UNTIL<sup>73</sup>. We consider, therefore, that it would be relevant to include this pattern of grammaticalization, so that it can be taken into account when approaching the evolution of the codification of the purpose in other languages.

(3) Consecutive-intensifier value. This is the second of the values of *hattā* with the source domain of SPATIOTEMPORAL LIMIT, illustrated by the example (4). In order to conceptualize the process of adoption of this new value, the semantic category of *force dynamics*<sup>74</sup> is very illustrative. In the consecutive-intensifier *hattā*, the initial trajectory not only has reached the limit, but has surpassed it, generating a consequence. In terms of force dynamics, we would say that a first force, the *Agonist*, is able to surpass the resistance applied by the second force, the *Antagonist*, and therefore, it originates a consequence. TIME is a source domain to generate elements that codify cognitively more complex notions, such as the consequence. As Heine and Kuteva point out, this pathway is “an instance of a more general process whereby spatial and temporal markers are grammaticalized in specific contexts to mark-ers of “logical” relations”<sup>75</sup>, relations such as cause-consequence. In Spanish there is the case of *luego* ‘afterwards’ > *luego* ‘therefore’<sup>76</sup>, from a temporal to a consecutive domain. In Arabic, this pathway could be outlined as follows:

(21) LIMIT > SPATIOTEMPORAL LIMIT > CONSECUTIVE-INTENSIFIER

The characteristic intensive value that *hattā* exhibits since the beginnings could be related to the fact that, in the dichotomy *hattā* ‘until’ / *‘ilā* ‘towards, until’, the first is the marked element. More research would be necessary, but it seems one of the most explanatory factors to account for the intensive, scalar and valorative values that *hattā* develops already in early stages.

(4) Scalar value. The scalar focus adverb ‘even’ also develops from the temporal ‘until’. This change of value, from temporal to scalar, is carried out through a metonymical process of transference of meaning.

<sup>73</sup> Heine and Kuteva, “World Lexicon”, p. 335.

<sup>74</sup> Talmy, “Force dynamics”.

<sup>75</sup> Heine and Kuteva, “World Lexicon”, p. 40.

<sup>76</sup> Garachana, “Los procesos de gramaticalización”, p. 163.

More concretely, it happens through the semantization of a conversational implicature, that is, of a meaning implicated in specific linguistic contexts<sup>77</sup>. This way, the limit of the trajectory, movement or time from the preposition and conjunction *ḥattā* ‘until’ changes into the limit of a scale (‘even’). And this scale, conceptualized vertically, includes a pragmatic classification of possibilities<sup>78</sup>; the clause introduced by ‘even’ is found, therefore, in the highest or lowest position of this scale, in one of its limits, and it is a more informative clause, as it is linked to contextual values such as the surprise or the counter-expectation<sup>79</sup>.

This kind of evolutionary pathway, in which a spatiotemporal conjunction or preposition transforms into a scalar focus adverb (UNTIL > EVEN), seems rather common, at least in Romance languages and Arabic. Concretely, we find it in Italian (*fino*), Portuguese (*até*), Catalan (*fins* > *fins i tot*), Romanian (*până* > *până și*) or Spanish (*hasta*), where these elements codify both the spatiotemporal limit and the scalar value. Surprisingly, it does not appear mentioned in the WLG: the authors only include one process in which UNTIL is the source domain: UNTIL > EQUATIVE COMPARATIVE<sup>80</sup>. The pathway of *ḥattā* until reaching the scalar value could be outlined as follows:

(22) LIMIT > SPATIOTEMPORAL LIMIT > SCALAR

(5) Temporal-disruptive value. Finally, this value of *ḥattā*, translated for ‘when (suddenly)’ (see example 6), also comes from the temporal value of the particle. The temporal *ḥattā* highlights the final part of a trajectory as well as its limit; the temporal-disruptive *ḥattā*, on its side, interposes a trajectory B before the first trajectory (A) reaches its end. Somehow, the temporal *ḥattā* codifies the arrival to a limit, and the disruptive *ḥattā* imposes a limit from which a new beginning is initiated. This specific pathway has not been found in the WLG; however, as it has been mentioned in section 4.1.2, we defend the change from a temporal to a disruptive value took place through a reanalysis. The first examples found in the corpus are bridging contexts: they have an ambiguous reading (see example 10), and they present always the follow-

<sup>77</sup> Grice, “Logic and Conversation”.

<sup>78</sup> Fuentes Rodríguez, “Enlaces extraoracionales”, p. 165.

<sup>79</sup> Rodríguez Rosique, “Pragmática y Gramática”, p. 302.

<sup>80</sup> Heine and Kuteva, “World Lexicon”, p. 326.

ing structure: ‘it did not take long *until/when*’. Both interpretations are possible, and when an author chooses the second reading (‘when’), he generates a more specific semantic and syntactic context that allows only the disruptive reading (see example 11). Both kinds of *hattā* operate in the temporal domain, and both codify an essential concept in this particle: the limit. The pathway could be described as follows:

$$(23) \quad \text{LIMIT} > \text{SPATIOTEMPORAL LIMIT} > \text{TEMPORAL-DISRUPTIVE}$$

## 5.2. Recapitulation

This cognitive approach has allowed us to develop a conceptual map that organizes all values of *hattā*, and to link the evolutionary pathways of this element with more general patterns of language change. At the same time, we have been able to identify some of the specificities of the history of this particle and, in general, of the processes of language change in Classical Arabic. We would like to mention three relevant observations about this approach:

(i) The starting point of our study has been the analysis of fragments extracted from the historical Arabic corpus HADC<sup>81</sup>, still unpublished. This methodology has allowed us to work with real examples of use, and to access texts spanning 1400 years. Therefore, we have been able to confirm that the appearance and increase of frequency of each kind of *hattā* correspond to their position in the conceptual map: the more at the bottom they are, the later they appeared, or the later their frequency increased. Not only the cognitive approach suggests this analysis, but the data from the corpus and the frequency of use corroborate it.

(ii) Another of the relevant aspects of this analysis is the fact that the values of *hattā* are more abstract, more cognitively complex and more valorative as we descend in the map, in line with the predictions of the grammaticalization framework. Indeed, the evolutionary pathway of *hattā* would have its origins in the referential and concrete word ‘limit’, ‘border’ (*ḥadd*); then, it would have passed to codify, in a more grammatical way (through a conjunction and a preposition) the notion of spatiotemporal limit. Thereafter *hattā* would adopt values cognitively more complex and more abstract, such as relations of cause-con-

<sup>81</sup> Al-Said, “A Corpus-based Historical”.

sequence (consecutive-intensifier, ‘to the point that’) and purpose (‘so that’). The particle would continue its trajectory adopting even more abstract meanings, and also more valorative and subjective, such as the scalar value (‘even’).

(iii) A typological contrast has been added to this cognitive approach: each of the semantic changes of *ḥattā* has been searched in other languages, mainly—but not only—through the work *Word Lexicon of Grammaticalization*. Therefore, it has been possible to strengthen some existing patterns and, at the same time, to broaden the scope of other pathways that had been found only in a specific language family. And, even more important, we have described evolutionary trajectories that were not included in the WLG, but that are fruitful not only in Arabic but also in languages from other language families.

## 6. Conclusions

This research began with a two-fold objective: first, to produce a methodological model for the study of language change in Classical Arabic; second, to analyse how this Semitic language evolved and changed, in particular concerning grammaticalization processes, through the example of the particle *ḥattā*. The main conclusions of this work can be summarized in five points:

(i) It has been provided a rigorous methodology to tackle, to the extent possible, the challenges presented by this kind of analysis. We have found and adapted the corpus HADC; data from historical periods barely taken into account in previous studies have been included; 30 variables have been analysed, many of them not considered in previous accounts. This methodological proposal also integrates a qualitative and quantitative analysis of an element throughout fourteen centuries, and provides a relational database that allows developing multiple analyses considering other of the gathered variables.

(ii) This diachronic approach shows that written Classical Arabic does in fact evolve: it has done so as long as it has existed, and not only semantically, but also morphosyntactically and pragmatically. *Ḥattā*, in the earliest period under investigation (the Quran, 7th century), only possessed a single fixed value, as a temporal limit (‘until’); the four remaining values of this particle in the sacred book were not

fixed<sup>82</sup>, and could in every example be subject to two different interpretations (the value as a temporal limit being acceptable in all cases). However, at the end of this journey, in the 20th century, we noted that all of the Quranic values of *ḥattā* were, in most examples, completely fixed; moreover, a new value was also present, the temporal-disruptive *ḥattā* ('when [suddenly]'), which goes back at least as far as the 9th century. Finally, we observed that the usage frequency of each value of *ḥattā* varies greatly over the centuries.

(iii) On the other hand, this study also shows that *fuṣḥā*, although it does evolve, does it in a more constrained and limited fashion. Indeed, the values of this particle found in the Quran are the same, and perhaps more importantly, nearly the only ones, which we see throughout its history. The constraints placed on language change in Classical Arabic are mainly due to the sacred nature of *fuṣḥā* as well as its intrinsic bond to the written form, which is not subject to the same tensions as orality. It is therefore essential to bear these constraints in mind when studying the evolution of this language.

(iv) Our research has also found that variables such as the geographic origin of authors, or the type of text in which changes are seen, are useful for this analysis. Specifically, it has been observed that the cases of *ḥattā* with an intensifier, scalar and subjective nature have been mainly developed in travel literature texts from al-Andalus, and more research would be necessary to expand on this first approach.

(v) Finally, one of the most remarkable features of language change in Classical Arabic is the marked *layering*, *persistence* and *divergence* that can be observed over the course of the evolution of *ḥattā*. These three principles were established by Hopper<sup>83</sup> in order to detect an ongoing process of grammaticalization; our analysis has shown that they are far more numerous and long-lasting in Classical Arabic than in other languages. Precisely for this reasons, these principles hold less significance. That is, examples of *layering*, *persistence* or *divergence* do not automatically imply that the observed change is active or ongoing.

The results of this research broaden our understanding of the evolutionary nature of Arabic and challenge some of our preconceptions

<sup>82</sup> Or not even present, such as the temporal-disruptive value 'when (suddenly)'.

<sup>83</sup> Hopper, "On some principles".



concerning this language<sup>84</sup>, which all too often prevent us from approaching it from a diachronic perspective. On the other hand, when comparing the history of *ḥattā* in Classical Arabic and in the Arabic dialects<sup>85</sup>, we see the evolution of this particle is far richer and more complex in the latter<sup>86</sup>. Therefore, this finding reveals that both *fuṣṣḥā* and the dialects should be taken into account when approaching Arabic from a diachronic perspective. Both expressions of this language, Classical Arabic and vernaculars, build up a system that must be examined together if one is to perform a rigorous and complete linguistic analysis of what is called Arabic.

Finally, we would like to mention that the conclusions of this work are not only related to the field of Arabic linguistics, but are also relevant to the field of historical linguistics overall, as they provide data about a language to which little study has been devoted. As Bisang<sup>87</sup> notes, grammaticalization processes are not homogeneous in every language and that is precisely why every theory must build on data from as many languages as possible, as well as incorporate their specific features. Otherwise, the explanatory capacity of the theoretical proposal is considerably limited and its universal value might be questionable.

## 7. References

- Al-Jallad, Ahmad, “The Etymology of *Ḥattā*”, in Nehmé and Al-Jallad (eds.), *To the Madbar and Back Again. Studies in the languages, archaeology, and cultures of Arabia dedicated to Michael C.A. Macdonald*, Leiden, Brill, 2018.
- Al-Jallad, Ahmad. *An outline of the grammar of the Safaitic inscriptions*, Leiden, Brill, 2015.
- Al-Said, Almoataz B., *A Corpus-based Historical Arabic Dictionary: Linguistic & Computational processing*, PhD dissertation, Cairo University, 2011.
- Ali, Abdullah Yusuf, *The Holy Quran: English Translation and Commentary*, Ware, Wordsworth Editions, 1934 (2001).

<sup>84</sup> Owens, “A linguistic history”.

<sup>85</sup> For a detailed analysis of *ḥattā* in the Arabic dialects, see Medea-García, “Gramaticalización y cambio lingüístico”, pp. 241-302.

<sup>86</sup> Vernacular languages, even when only considering the urban dialect of each country’s main city, account for three times as many values of *ḥattā* as Classical Arabic: eighteen in total (see Medea-García, “Gramaticalización y cambio lingüístico”, p. 280).

<sup>87</sup> Bisang, “Grammaticalization without coevolution”.



- Arberry, Arthur John, *The Koran Interpreted: A Translation*, New York, Touchstone, 1955 (1996).
- Bisang, Walter, "Grammaticalization without coevolution of form and meaning: The case of tense—aspect—modality in East and mainland Southeast Asia", in W. Bisang, N. P. Himmelmann, and B. Wiemer (eds.), *What makes Grammaticalization?: A Look from its Fringes and its Components*, Berlin, Walter de Gruyter, 2004, pp. 109-138.
- Bybee, Joan L., "Cognitive processes in grammaticalization", in Michael Tomasello (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, Hillsdale, NJ, Lawrence Erlbaum, 2003, pp. 145-167.
- Chejne, Anwar G., *The Arabic language: Its role in history*, Minnesota, University of Minnesota Press, 1969.
- Corriente, Federico, Christophe Pereira, and Ángeles Vicente, (eds.), *Dictionnaire du faisceau dialectal arabe andalou: Perspectives phraséologiques et étymologiques*, Berlin, Walter de Gruyter, 2017.
- Cuenca, Maria Josep, "La gramaticalización", in I. Ibarretxe-Antuñano and J. Valenzuela (eds.), *Linguística cognitiva*, Barcelona, Anthropos, 2012, pp. 281-334.
- Diewald, Gabriele, "A model for relevant types of contexts in grammaticalization", *Typological Studies in Language*, 49 (2002), pp. 103-120.
- Diewald, Gabriele, "Grammaticalization and pragmaticalization", in Heiko Narrog and Bernd Heine (eds.), *The Oxford Handbook of Grammaticalization*, Oxford, Oxford University Press, 2011, pp. 450-461.
- Diewald, Gabriele, "Pragmaticalization (defined) as grammaticalization of discourse functions", *Linguistics*, 49/2, (2011), pp. 365-390.
- Dozy, Reinhart Pieter Anne. *Supplément aux dictionnaires arabes*, Leiden, Brill, vol. 1, 1881.
- Esseesy, Mohssen, *Grammaticalization of Arabic prepositions and subordinators: A corpus-based study*, Leiden, Brill, 2010.
- Evans, Nicholas and Wilkins, David, "In the mind's ear: The semantic extensions of perception verbs in Australian languages", *Language*, 76 (2000), pp. 546-592.
- Ferrando, Ignacio, *Introducción a la historia de la lengua árabe: nuevas perspectivas*, Zaragoza, Universidad de Zaragoza, 2001.
- Fuentes Rodríguez, Catalina, *Enlaces extraoracionales*, Madrid, Ediciones Alfaro, 1987.
- Garachana, Mar, "Los procesos de gramaticalización", *Moenia*, 5 (1999), pp. 155-172.
- Grice, Herbert Paul, "Logic and Conversation", in P. Cole, y J. L. Morgan (eds.), *Syntax and Semantics*, New York, Academic Press, 1975, pp. 41-58.
- Haeri, Niloofar. *Sacred language, ordinary people: Dilemmas of culture and politics in Egypt*, Springer, 2003.

- Heine, Bernd, "On the role of context in grammaticalization", *Typological Studies in Language*, 49 (2002), pp. 83-102.
- Heine, Bernd, and Tania Kuteva. *World lexicon of grammaticalization*, Cambridge, Cambridge University Press, 2002.
- Hopper, Paul J., "On some principles of grammaticization", in Elizabeth Closs Traugott and Bernd Heine (eds.), *Approaches to grammaticalization*, Amsterdam, John Benjamins, vol. 1, 1991, pp. 17-35.
- Ibn Hišām, Ġamāl al-Dīn Abū Muḥammad 'Abdallāh b. Yūsuf b. Aḥmad b. 'Abdallāh, *Muġnī al-labīb 'an kutub al-'a'ārīb*, Māzin al-Mubārak and Muḥammad 'Alī Ḥamd Allāh (eds.), Beirut, Dār al-Fikr, 1995.
- Ibn Manẓūr, Muḥammad Ibn Mukarram, *Lisān al-'arab*, Beirut, Dar Ṣādir, 15 Vols, 1290 (1955).
- Irbillī, Alī Ibn Muḥammad, *Kitāb Ġawāhir al-'adab fī ma'rifāt al-'arab*, Cairo, Maṭba'at Wādī al-Nīl al-Miṣrīya, 1294/1877.
- Kabatek, Johannes. "Tradiciones discursivas jurídicas y elaboración lingüística en la España medieval", *Cahiers d'études hispaniques medievales* 27/1 (2004), pp. 249-261.
- Kabatek, Johannes. "Tradiciones discursivas y cambio lingüístico." *Lexis* 29/2 (2005), pp.151-177.
- Khan, Mohsin, *The Noble Quran in the English Language*, Saudi Arabia, King Fahd Printing Complex, 1996.
- Lakoff, George and Johnson, Mark, *Metaphors we live by*, Chicago, University of Chicago Press, 1980.
- Lane, Edward William, *An Arabic-English Lexicon: Derived from the Best and the Most Copious Eastern Sources*, London, Williams and Norgate, 1865-1893.
- Langer, Michael, "Conjunctions", in Kees Versteegh (ed.), *Encyclopedia of Arabic Language and Linguistics*, Leiden, Brill, pp. 467-470.
- Maldonado, Ricardo, "La gramática cognitiva", in I. Ibarretxe-Antuñano and J. Valenzuela (eds.), *Lingüística cognitiva*, Barcelona, Anthropos, 2012, pp. 213-248.
- Medea-García, L., and Al-Said, Almoataz, "The Historical Arabic Dictionary Corpus and its Suitability for a Grammaticalization Approach: A Case Study", communication presented at the *V International Conference: Grammar and Corpora*, Warsaw, 2014.
- Medea-García, L., "Gramaticalización y cambio lingüístico en árabe. El caso de ḥattā en lengua clásica y dialectal", PhD dissertation, Barcelona, Universitat Autònoma de Barcelona, 2017, retrieved from <https://www.tdx.cat/bitstream/handle/10803/459246/lmg11de1.pdf?sequence=1&isAllowed=y>.
- Melis, Chantal, Marcela Flores, and Sergio Bogard. "La historia del español. Propuesta de un tercer período evolutivo." *Nueva Revista de Filología Hispánica* 51/1 (2003), pp. 1-56.
- Owens, Jonathan. *A linguistic history of Arabic*, Oxford, Oxford University Press, 2006.

- Pickthall, Mohammed Marmaduke, *The Meaning of the Glorious Koran*, New York, Plume, 1930 (1997).
- Pons Bordería, Salvador. "El siglo xx como diacronía: intuición y comprobación en el caso de 'o sea'", *RILCE*, 30/3 (2014), pp. 985-1016.
- Rendsburg, Gary A., Aaron D. Rubin, and John Huehnergard. "A proper view of Arabic, Semitic, and more", *Journal of the American Oriental Society*, 128/3 (2008), pp. 533-541.
- Rodríguez Rosique, Susana, *Pragmática y Gramática: condicionales concesivas en español*, Frankfurt, Peter Lang, 2008.
- Sadan, Arik, *The subjunctive mood in Arabic grammatical thought*, Leiden, Brill, 2012.
- Sahih International, *The Quran: Arabic Text with Corresponding English Meanings*, Saudi Arabia, Abul Qasim Publishing House, 1997.
- Sarwar, Muhammad, *The Holy Qur'an: The Arabic Text and English Translation*, London, Elmhurst, 1981.
- Talmy, Leonard, "Force dynamics in language and cognition", *Cognitive science*, 12/1, (1988), pp. 49-100.
- Traugott, Elizabeth Closs and König, Ekkehard, "The semantics-pragmatics of grammaticalization revisited", in Elizabeth Closs Traugott and Bernd Heine (eds.), *Approaches to grammaticalization*, Amsterdam, John Benjamins, vol. 1, 1991, pp. 189-218.
- Traugott, Elizabeth Closs, "Meaning-change in the development of grammatical markers", *Language sciences*, 2/1, (1980), pp. 44-61.
- Traugott, Elizabeth Closs, "On the rise of epistemic meanings in English: An example of subjectification in semantic change", *Language*, 65/1, (1989), pp. 31-55.
- Traugott, Elizabeth Closs, "From less to more situated in language: the unidirectionality of semantic change", in *5th International Conference on English Historical Linguistics*, Amsterdam-Philadelphia, John Benjamins Publishing Company, 1990, pp. 497-517.
- Traugott, Elizabeth Closs, "Subjectification in grammaticalization", in Dieter Stein and Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, Cambridge, Cambridge University Press, 1995, pp. 31-54.
- Traugott, Elizabeth Closs, "Semantic change: An overview", *Glott international*, 2/9-10, (1996). pp. 3-7.
- Traugott, Elizabeth Closs, "(Inter)subjectivity and (inter)subjectification: A reassessment", in Kristin Davidse, Lieven Vandelanotte, and Hubert Cuyckens (eds.), *Subjectification, intersubjectification and grammaticalization*, Berlin, Walter de Gruyter, 2010, pp. 29-71.
- Versteegh, Kees, *The Arabic Language*, Edinburgh, Edinburgh University Press, 2014.
- Visconti, Jacqueline. "On the origins of scalar particles in Italian", *Journal of Historical Pragmatics*, 6/2 (2005), pp. 237-261.

Wright, William, *A grammar of the Arabic language*, New Delhi, Anmol Publications, 1999 [First ed.1859-1962].

*Recibido:* 02/09/2018

*Aceptado:* 15/11/2018